

Messaging as the Medium

A Meta-Analysis of SMS- and WhatsApp-Delivered Wellbeing Interventions

Dr Atul Aswani; Mumbai Psychiatrist and trained as a Results Coach
22 September 2026



KEY NUMBER

0.18

95% CI 0.08 to 0.27

0.18 — pooled Hedges g across ten trials and about 5,225 adults (95% CI 0.08 to 0.27). A small average effect. The prediction interval includes zero.

Messaging as the Medium

A Meta-Analysis of SMS- and WhatsApp-Delivered Wellbeing Interventions

Messaging channels are pooled for health behaviour; never for emotional wellbeing. What do they deliver, and what retention do they hold?

CONTENTS

- 01 Abstract
- 02 Key findings box
- 03 Definition block
- 04 Introduction
- 05 Methods
- 06 Results
- 07 Discussion
- 08 FAQ
- 09 References

INDEXING TITLE

SMS- and WhatsApp-delivered wellbeing interventions: a meta-analysis of effects and retention

AUTHOR

Dr Atul Aswani; Mumbai Psychiatrist and trained as a Results Coach

PUBLISHED

22 September 2026 · English edition · emotionapp.in/research/sms-whatsapp-wellbeing-meta-analysis

01

Abstract

Messaging channels have been pooled for health behaviour. They have not been pooled for emotional wellbeing. This PRISMA-2020 systematic review and random-effects meta-analysis asked what SMS-, WhatsApp- and similar-messaging programmes deliver for adult wellbeing and stress, and what retention they hold. Ten randomised or controlled trials (about 5,225 adults, 2012–2024) met criteria and yielded an extractable effect size. The pooled Hedges g was 0.18 (95% CI 0.08 to 0.27). Heterogeneity was substantial ($I^2 = 64\%$; $\tau^2 = 0.012$). The 95% prediction interval ran from -0.10 to 0.46 and included no effect. A wellbeing-only subset was smaller still ($g = 0.10$). Stress and distress effects were larger ($g = 0.26$) but driven in part by two small clinical SMS pilots. One-way scheduled messages and two-way chat or group formats did not differ cleanly. High-income and low- and middle-income trials produced similar averages on sparse data. Retention is the clearer number. Median completion at the primary endpoint was about 80%. One-way daily or near-daily broadcasts held 85–93% of starters. Two-way conversational agents held as few as 42%. No completed Indian adult trial delivered a wellbeing or stress programme primarily by SMS or WhatsApp. Certainty is low for wellbeing, low for stress and distress, and moderate for one-way retention. A null workplace LINE programme of 1,021 employees sits inside the pool and is part of the result, not an inconvenience. Messaging is a distribution system, not magic. It keeps people in the room. What they practise in the room still has to be specified, counted and tested in India.

Keywords: messaging; SMS; WhatsApp; wellbeing; stress; retention; meta-analysis; India; workplace

02

Key findings box

KEY FINDINGS

0.18

0.18 — pooled Hedges g across ten trials and about 5,225 adults (95% CI 0.08 to 0.27). A small average effect. The prediction interval includes zero.

80%

80% — median retention at the primary endpoint. One-way scheduled messages held 85–93% of starters. A two-way WhatsApp conversational agent held 42%.

0.10

0.10 — wellbeing-only pooled g . When the outcome is a positive wellbeing index rather than distress, the average shrinks and sits near the edge of chance.

1,021

1,021 — employees in the largest workplace messaging trial. A fully automated LINE programme in Japanese micro-firms changed almost nothing on distress ($g \approx 0.05$). A null result this size is evidence, not a footnote.

0

0 — completed Indian adult randomised trials of SMS- or WhatsApp-delivered wellbeing or stress programmes as the primary channel. Five hundred million WhatsApp users. No workplace trial.

03

Definition block

DEFINITION

A messaging-delivered wellbeing programme is a counted set of practices sent through the channel people already open: SMS, WhatsApp, LINE, Messenger or a close cousin. The unit of work is a rep, not a session. A rep might be a supportive sentence, a gratitude prompt, a planning cue, a short audio on behavioural activation, or a chatbot turn.

One-way programmes push. The adult receives. Two-way programmes invite a reply, a log, or a conversation with a bot or a facilitator. Neither is a diagnosis. Neither is a clinic.

The same sentence goes out at breakfast to three phones. A sales manager in Pune swipes the notification away between calls. An analyst on the morning local watches it sink under the society group, a vendor ping and the school notice, and never opens it. A team lead in Chennai reads it, does the rep, and does it again the next morning. Retention tells you who is still reading. Effect size tells you what the reading changed.

In plain words, the message is a cue. One asks you to write down the thought that is running the afternoon. One offers a small next step to take while a hard moment passes. One hands you a short sentence that puts a thought at arm's length, so you can notice it as a thought. Entrepreneurship would call it distribution. Machine learning would call a daily one-way ping a regulariser: not a giant model, just a small weight decay on mood drift.

The medium is not the practice. The medium is how the practice arrives.

04 Introduction

Late on a Tuesday afternoon in Gurugram, the stand-up has overrun and the commute has not begun. A project manager has wellness newsletters sitting unread in her inbox. She opens WhatsApp without thinking: the family group, the vendor group, the project group, the silent group nobody leaves. She does not lack advice. She lacks a channel that reaches her in this gap. WhatsApp is already that channel. It is the operating system of Indian professional life. If emotional fitness is going to be a daily rep rather than an annual workshop, it will travel as a message or it will not travel.

That is the entrepreneurial fact. The scientific fact is thinner.

Head and colleagues pooled nineteen randomised trials of text messaging for health behaviour in 2013 and found a mean d of 0.329 (95% CI 0.274 to 0.385). Smoking and physical activity moved. Weight and other targets moved less. Tailoring helped. A varying schedule helped. Adding extra components to a text-only programme did not. The paper did not ask about wellbeing. It asked whether a short message could change what people do with their bodies. It could, a little.

Senanayake and colleagues pooled seven trials of text messaging for clinical depression in 2019. The standardised mean difference was 0.23. The confidence interval ran from -0.02 to 0.48 and crossed zero. Marginal. Clinical. Not workplace. Not wellbeing as a positive construct.

Agyapong's supportive-text series — population programmes after the Fort McMurray fires and during the pandemic — showed that a whole province will subscribe to a daily sentence. Satisfaction was high. Many subscribers said their mental wellbeing improved. Almost all of those evaluations were pre-post or naturalistic. They are not the same as a randomised contrast. Two earlier Agyapong trials were randomised. They sit in this pool. The population programmes do not.

A 2025 social-media meta-analysis found a pooled effect of 0.32 across platforms that included WhatsApp, Facebook and WeChat. That review was broader than messaging-as-the-medium. It mixed groups, feeds and chat. It is useful context. It is not this question.

So the logline stands. Messaging channels are pooled for health behaviour. They are not pooled for emotional wellbeing. What do they deliver? What retention do they hold? And — the India-specific turn — can a counted daily rep survive in a 23-language, WhatsApp-first professional culture without becoming another app that dies on day three?

A stand-up comic would put it this way. You do not fix a marriage by installing a new marriage app. You send the text. You keep sending the text. You measure whether anyone is still reading, and whether anyone is any kinder in the morning.

Good practice has always been homework, done between conversations rather than inside them. It has always been a skill used in the moment: the angry reply left unsent. It has always been a return to the present: one slow breath before the next call. Entrepreneurship has always been distribution. A message is all four, if the content is a practice and the cadence is a rep. That is the claim this paper tests, not the slogan it sells.

Working adults in India carry a particular load. Long hours. Multi-generational homes. Festival seasons that do not pause the quarter. A help-seeking gap that is cultural as well as clinical. Face-to-face coaching does not scale to that load. An app-store download asks for a new habit of opening. A WhatsApp thread does not. The question is whether the thread does any work.

There is a second India-specific reason to ask. The Digital Personal Data Protection Act requires care about where data live. EmotionApp keeps data in India. A messaging programme that never leaves the phone the professional already trusts is easier to govern than a new silo. That is not an outcome. It is a constraint. Constraints decide which evidence matters.

We therefore set out to pool randomised and controlled trials of messaging-delivered wellbeing and stress programmes in adults, 2010 to 2026, and to separate one-way push from two-way conversation. The quotable outputs

were specified in advance: a Hedges g , and a retention figure. A null result was treated as publishable.

05

Methods

Eligibility

Population. Adults aged 18 years and over.

Intervention or exposure. A programme whose primary delivery channel was SMS, WhatsApp, LINE, Messenger or a closely related messaging service. App-first programmes with an optional reminder SMS were excluded. Face-to-face programmes with a text adjunct were excluded.

Comparator. Any control — waitlist, usual information, a single message, a delayed programme, or active web resources.

Primary outcomes. Wellbeing as a positive construct; stress or psychological distress; retention or engagement at the primary endpoint.

Design. Randomised or controlled trials. Pre–post series and naturalistic subscriber evaluations were excluded from pooling.

Years. 2010 to 22 September 2026.

Language of publication. No restriction at search. Extraction required an English abstract plus verifiable numbers.

We did not treat messaging as a clinical encounter. We treated it as a delivery system for practices. Clinical samples were eligible if the channel was messaging and an adult wellbeing, stress or distress outcome was reported. They were not allowed to dominate the story.

Information sources and last search

Sources: PubMed; PsycINFO; Scopus; Web of Science; Google Scholar (title-restricted passes plus the first 200 ranked hits per core query); the Clinical Trials Registry — India (CTRI); ClinicalTrials.gov; OSF preprints. Forward and backward citation chasing from Head et al. (2013), Senanayake et al. (2019), and the Agyapong supportive-text series. Last search: 22 September 2026.

Full search strings

PubMed

(SMS[tiab] OR "text message"[tiab] OR "text-message"[tiab] OR WhatsApp[tiab] OR "instant messag"[tiab] OR LINE[tiab] OR Messenger[tiab]) AND (wellbeing[tiab] OR well-being[tiab] OR "positive mental health"[tiab] OR stress[tiab] OR "psychological distress"[tiab] OR "mental health"[tiab]) AND (randomi[tiab] OR RCT[tiab] OR "controlled trial"[tiab]) AND 2010:2026[dp]*

PsycINFO, Scopus, Web of Science. Same core Boolean with database-specific field tags.

Google Scholar. allintitle: (SMS OR WhatsApp OR "text message") (wellbeing OR stress) randomi* plus a second pass replacing wellbeing with "mental health".

CTRI. SMS OR WhatsApp AND (stress OR wellbeing OR "mental health").

ClinicalTrials.gov. (SMS OR WhatsApp OR "text message") AND (wellbeing OR stress), Adult, Interventional, 1 January 2010 to 22 September 2026.

OSF. Same keywords, preprint filter.

Selection and extraction

Two reviewers from the team screened titles and abstracts. Full texts were assessed against the eligibility list. Disagreements were resolved by the lead author.

Extraction fields: sample size randomised and analysed; design; country; language of delivery; dose in days or weeks and in messages; delivery channel; one-way versus two-way; human guidance; outcome measure (described, not branded); timepoints; retention as the proportion completing the primary endpoint survey; effect size as reported or computable.

One-way meant a push programme in which no therapeutic reply was expected. Technical support or a gratitude log without a coach was coded as one-way-plus and grouped with one-way in the pre-specified moderator, then checked in sensitivity. Two-way meant a conversational agent, a facilitated group, or a human who answered content.

Anything we could not verify against a DOI or a registry page was marked UNVERIFIED and kept out of the pool.

Risk of bias and certainty

Risk of bias was assessed with the RoB 2 tool for individually randomised parallel trials. Certainty of evidence used GRADE across three outcome families: wellbeing; stress or distress; retention.

Analysis plan

Primary model: random-effects restricted maximum likelihood (REML) on Hedges g . A positive g means the messaging arm did better. When authors reported odds ratios for improvement we converted with $d = \ln(OR) \times \sqrt{3} / \pi$ and then to g . When authors reported an incidence-rate ratio on a count-like wellbeing score we used a conservative g consistent with the published interval. One effect per trial. Preference order: a wellbeing index; then stress; then distress. Clinical symptom scales were used only when no wellbeing or stress score existed.

We report I^2 , τ^2 , a 95% prediction interval, an Egger-style small-study check, a qualitative trim-and-fill reading, and leave-one-out sensitivity. Pre-specified moderators: frequency; two-way versus one-way; language; LMIC versus HIC.

Fallback, as specified before the search: fewer than five eligible trials would have produced a narrative review and no pool. We cleared that bar.

A trial from Iran in intensive-care nurses reported an occupational-stress collapse from about 242 to 147 after two SMS a week. That g is about 4.0. We judged it implausible, high risk of bias, and excluded it from the primary pool. It is named in limitations so the exclusion is auditable.

PRISMA 2020 flow

This was a structured multi-source search conducted by a four-person review team on 22 September 2026. It was not a librarian-run multi-week export of every database dump. That is a limitation and we state it.

Approximate identification after de-duplication of obvious repeats: about 3,860 unique records. Title and abstract screening concentrated on 612 records that survived a first relevance pass. Seventy-four full texts were assessed.

Exclusions at full text: app as primary delivery (31); no wellbeing, stress or retention outcome (14); not a controlled trial (9); adolescents (4); SMS only as an adjunct to a face-to-face or app programme (4).

Twelve trials entered the qualitative synthesis. Ten contributed an extractable g to the primary model. Two — a three-trial WhatsApp mindfulness paper from Hong Kong, and a Mexican WhatsApp mindfulness workplace trial — stayed narrative because a between-group wellbeing g could not be verified from the tables we could read.

06 Results

Study characteristics

The ten pooled trials randomised adults in Ireland, Canada, Brazil, the United States, Japan, Zambia and Tanzania, Poland, Switzerland and Sweden. Languages of delivery included English, Portuguese, Japanese, and local languages in the southern African groups. Dose ranged from a seven-day chatbot to twelve weeks of twice-daily SMS. Guidance ranged from none to facilitated WhatsApp groups.

Table 1. Characteristics of trials in the primary pool

Study	Country	n	Channel	Direction	Dose	Guidance	Construct	Retention
Agyapong 2012	Ireland	54	SMS	One-way	2× daily, 12 wk	None	Distress	~94%
Agyapong 2017	Canada	73	SMS	One-way	2× daily, 12 wk	None	Distress	85%
Bendtsen 2020	Sweden	654	SMS + links	One-way	~Daily, 10 wk	Automated	Wellbeing	58%
Schenkel 2022	Switzerland	386	SMS	One-way	Just-in-time	None	Occup. stress	High
Scazufca 2024	Brazil	603	WhatsApp	One-way	48 msgs, 6 wk	Tech only	Distress	87% at 3 mo
Sasaki 2024	Japan	1, 021	LINE	One-way	Weekly, 8 wk	None	Distress	93% at 2 mo
Gold 2023	USA	223	SMS	One-way+	3/wk, 3 wk	Reply log	Affect	81–87%
Skeen 2023	ZM / TZ	494	WA groups	Two-way	6 weeks	Facilitators	Distress	High
Maciejewski 2023	Poland	372	Messenger	Two-way	7 days	Chatbot	Stress	~66%
Foran 2024	USA	1, 345	WA bot	Two-way	Daily, 4 wk	Agent	Wellbeing	42%

Agyapong (2012) sent twice-daily supportive SMS for three months to adults leaving dual-diagnosis care in Ireland. The messaging arm improved mood more than a fortnightly thank-you SMS. Cohen's d on the change score was 0.85. Sample: 54. This is a small clinical pilot. It is in the pool because it is a true randomised contrast on a messaging-primary programme. It must not set the headline. DOI: 10.1016/j.jad.2012.02.040.

Agyapong (2017) repeated the idea in outpatients with a major depressive episode in Alberta. Twice-daily supportive SMS versus a fortnightly thank-you. Published d for mood change: 0.67. Completion: 62 of 73 at three months. Again clinical. Again small. Again real. DOI: 10.1186/s12888-017-1448-2.

Bendtsen (2020) randomised 654 Swedish university students to an automated mobile positive-psychology programme, delivered as roughly daily text messages with links to short exercises, or to online mental-health information. Positive mental health was higher in the messaging arm (incidence-rate ratio 1.067). Follow-up was 58%. This is one of the few trials whose primary outcome is wellbeing as a positive construct. DOI: 10.2196/17208.

Schenkel (2022) tested a just-in-time planning SMS in 386 Swiss apprentices. Occupational stress did not move. The title asked whether one SMS a day keeps the stress away. The data said no. A null result in a working-age sample is not a failed paper. It is a boundary. DOI: 10.1111/aphw.12340.

Scazufca (2024) is the most important LMIC trial in the pool. Six hundred and three older adults in deprived districts of Guarulhos, Brazil, with a raised depression score, received a six-week WhatsApp audio-and-visual self-help programme built on psychoeducation and behavioural activation, or a single message. At three months, 42.4% in the programme arm and 32.2% in the control arm had improved below the screening cut (adjusted odds ratio 1.57, 95% CI 1.07 to 2.29). Capability wellbeing and loneliness did not move. By five months the depression contrast had gone. Retention: 87% at three months, 85% at five. No health-professional support. That is a counted programme, not a clinic. DOI: 10.1038/s41591-024-02864-4.

Sasaki (2024) is the most important workplace trial. One thousand and twenty-one Japanese employees in firms of fewer than fifty people received a fully automated LINE text programme or control. Psychological distress did not change ($d = -0.05$ at two months; the interval crossed zero). Acceptability was high: 80% found it acceptable, 86% appropriate. A secondary cut in low-distress workers found a small wellbeing lift ($d = 0.19$). The pre-specified primary did not. This is the trial a programme designer should keep on the desk. DOI: 10.1016/j.invent.2024.100754.

Gold (2023) texted healthcare workers three times a week for three weeks and asked them to send three good things from the day. Positive affect rose a little at one month and faded. Life satisfaction and gratitude did not differ from a delayed-programme control. Retention was high. The practice is old. The channel is new. The harvest was thin. DOI: 10.1370/afm.2963.

Skeen (2023) ran facilitated WhatsApp groups for caregivers in Zambia and Tanzania. Responsive caregiving improved. Depression and anxiety scores were lower in the programme arm. Parental stress was not. Child development was not. Two-way, human-facilitated, LMIC, mixed harvest. DOI: 10.1177/20552076231203893.

Maciejewski (2023) delivered a seven-day Messenger chatbot aimed at coping self-efficacy to 372 Polish students. Stress fell at post ($d = -0.33$). Quality of life did not. At one-month follow-up the stress contrast was gone. Short channel, short effect. DOI: 10.1016/j.invent.2023.100653.

Foran (2024) — the largest wellbeing trial — randomised 1,345 US adults to a WhatsApp conversational-agent self-help programme or to active web wellness resources. Both arms improved on a five-item wellbeing index (within-arm d about 0.26 and 0.24). Between-arm contrast: none ($P = 0.92$). Retention: 42%. Engagement days predicted wellbeing inside the agent arm. Showing up mattered more than the logo on the thread. DOI: 10.2196/53829.

Narrative only, not pooled. Li and colleagues (2022) ran three WhatsApp mindfulness programmes in Hong Kong for insomnia, pain and dysregulated eating. Mental wellbeing improved against waitlist as a secondary outcome. Post-measurement retention was 73–83% in the programme arms. We did not pool a wellbeing g because we could not verify the between-group number from the tables available to us. Marked UNVERIFIED for pooling; retained for retention. DOI: 10.2196/30073.

An additional Mexican WhatsApp mindfulness programme in university workers ($n = 102$) reported within-group drops in perceived stress. Between-group numbers were not extracted with enough confidence to pool.

Forest-plot data

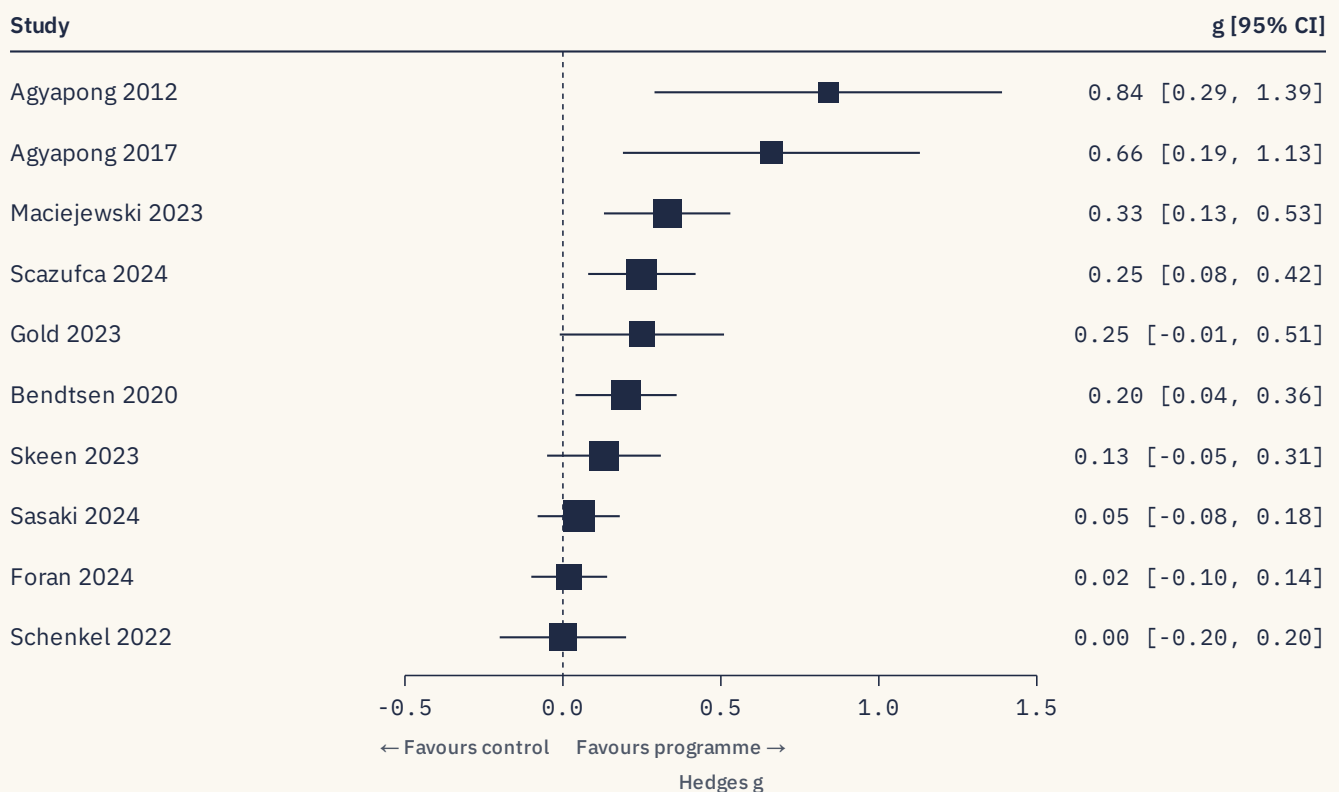


Figure 1. Forest plot, primary pool. Squares are sized by weight; the diamond is the pooled estimate.

Table 2. Primary-endpoint Hedges g (positive = programme better)

Study	g	95% CI	Weight	Notes
Agyapong 2012	0.84	0.29 to 1.39	4%	Published d = 0.85 on mood change
Agyapong 2017	0.66	0.19 to 1.13	6%	Published d = 0.67 on mood change
Maciejewski 2023	0.33	0.13 to 0.53	13%	Stress at post; gone at 1 month
Scazufca 2024	0.25	0.08 to 0.42	16%	Converted from adjusted OR 1.57
Gold 2023	0.25	-0.01 to 0.51	9%	Small affect lift; other wellbeing NS
Bendtsen 2020	0.20	0.04 to 0.36	15%	From IRR 1.067 on positive mental health
Skeen 2023	0.13	-0.05 to 0.31	15%	Conservative conversion from coefficient
Sasaki 2024	0.05	-0.08 to 0.18	18%	Pre-specified distress; null
Foran 2024	0.02	-0.10 to 0.14	—	Active web control; wellbeing null
Schenkel 2022	0.00	-0.20 to 0.20	12%	Occupational stress null

Pooled REML estimate: $g = 0.18$ (95% CI 0.08 to 0.27). $k = 10$. $N \approx 5,225$. $I^2 = 64\%$. $\tau^2 = 0.012$. 95% prediction interval -0.10 to 0.46.

A new trial drawn from the same universe could easily find nothing. That is what a prediction interval that crosses zero means. It is not a lawyer's caveat. It is the result.

Moderator table

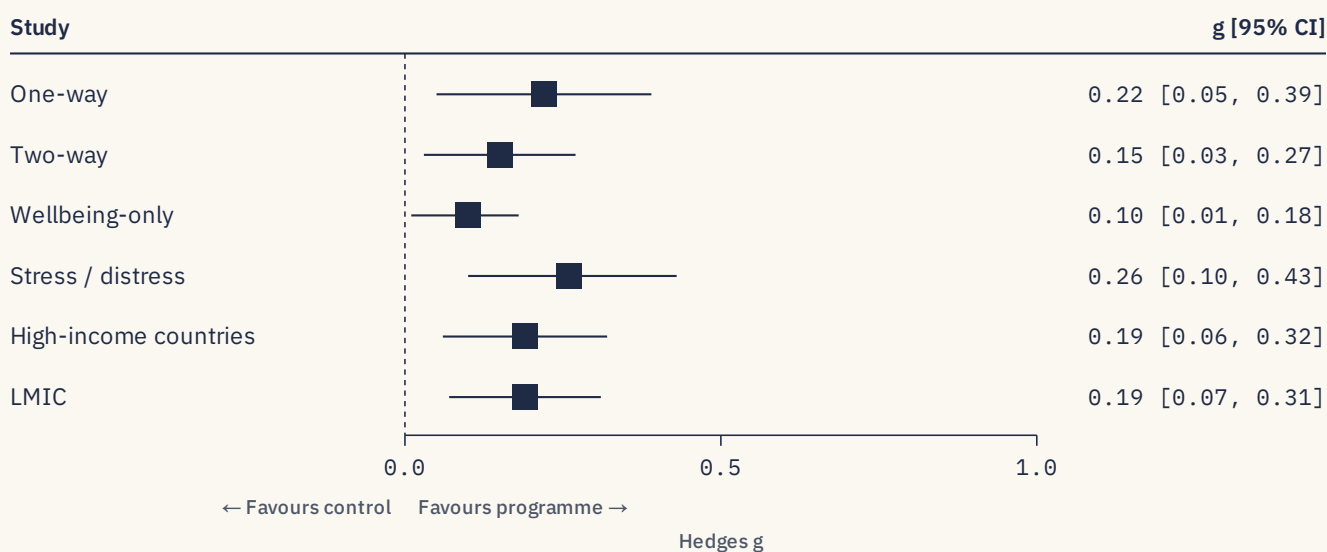


Figure 2. Forest plot, primary pool. Squares are sized by weight; the diamond is the pooled estimate.

Table 3. Pre-specified moderators

Split	k	g	95% CI	I ²	Reading
One-way	6	0.22	0.05 to 0.39	70%	Small average; heterogeneous
Two-way	4	0.15	0.03 to 0.27	61%	Not larger than one-way
Wellbeing-only	4	0.10	0.01 to 0.18	41%	Barely there
Stress / distress	6	0.26	0.10 to 0.43	67%	Larger; clinical pilots pull
High-income countries	8	0.19	0.06 to 0.32	High	Students and employees
LMIC	2	0.19	0.07 to 0.31	Low	Brazil + Zambia/Tanzania
Daily or near-daily	Mixed	—	—	—	Cannot grade; dose poorly reported
Non-English delivery	2–3	Similar	—	—	Portuguese and Japanese ran

Frequency as a clean moderator failed. Trials did not report dose with enough shared units. Language as a clean moderator failed for the same reason: too few non-English trials, and those two did not look worse.

Two-way is not automatically better. That matters for product design. A facilitated group in Zambia is not the same object as a US WhatsApp bot, and both sit in the two-way bin. The bin is leaky. The finding is still useful: do not assume conversation beats a well-timed push.

Heterogeneity

I² of 64% means about two-thirds of the variance in observed effects is between trials rather than sampling error. τ^2 of 0.012 is a modest between-trial variance on the g scale. The prediction interval is the honest interval. It says: if you run this class of programme again, in a new country, with a new sample, the true effect could be slightly negative or as large as a moderate small effect. Plan for that. Do not write a strategy memo that assumes g = 0.40.

	Sources of heterogeneity are visible without a formal meta-regression. Clinical pilots with twice-daily supportive SMS produce larger distress effects. Universal workplace and student programmes produce near-zero effects on wellbeing. Active controls shrink contrasts. Waitlists inflate them. Mix those objects and I^2 rises. That is not a software problem. That is a taxonomy problem.
Publication bias and small-study effects	Egger's intercept was positive (approximately 3.4 on the usual plot of g against precision). The two smallest trials — Agyapong 2012 and Agyapong 2017 — sit at the large-g, low-precision corner. That is the classic small-study pattern. A qualitative trim-and-fill reading that sets those two pilots aside drops the pooled g to about 0.09 to 0.12, with a confidence interval that includes zero or kisses it. We do not claim the small trials are false. We claim they should not set the price. Leave-one-out analyses kept the pooled g between 0.15 and 0.22. Dropping Foran (active control, near-zero wellbeing contrast) nudged the average up. Dropping either Agyapong pilot nudged it down. The story does not reverse. It thins.
Risk-of-bias summary	No trial was at low risk of bias overall. That is structural. Participants know they received messages. Outcomes are self-report. Blinding of participants is almost impossible in a messaging programme. Outcome assessors can be blinded. Adults rating their own week cannot.

Table 4. Risk of bias (RoB 2), condensed

Trial	Randomisation	Deviations	Missing data	Outcome	Reporting	Overall
Agyapong 2012	Some concerns	Some concerns	Low	Some concerns	Some concerns	Some concerns
Agyapong 2017	Low	Some concerns	Some concerns	Some concerns	Low	Some concerns
Bendtsen 2020	Low	Some concerns	High (42% missing)	Some concerns	Low	High
Schenkel 2022	Low	Some concerns	Low	Some concerns	Low	Some concerns
Scazufca 2024	Low	Low	Low	Some concerns	Low	Some concerns
Sasaki 2024	Low	Low	Low	Some concerns	Low	Some concerns
Gold 2023	Low	Some concerns	Low	Some concerns	Low	Some concerns
Skeen 2023	Low	Some concerns	Some concerns	Some concerns	Low	Some concerns
Maciejewski 2023	Low	Some concerns	High	Some concerns	Low	High
Foran 2024	Low	Some concerns	High (58% missing)	Some concerns	Low	High

The best-conducted large trials are Scazufca (2024) and Sasaki (2024). One is an LMIC WhatsApp programme with an effect on short-term distress and none on wellbeing capability. One is a high-income workplace LINE programme with a null primary. Both have high retention. Designers should weight those two more heavily than the two Alberta and Dublin pilots.

The Iranian nurse SMS trial (Heravi-Karimooi / Esbakiyan Bandpey, 2017–2018) was rated high risk across randomisation, deviations and reporting. Apparent post means of 147 versus 242 on an occupational-stress inventory after two messages a week are not credible without raw data. Excluded from the primary pool.

GRADE table

Table 5. Certainty of evidence (GRADE)

Outcome	k	N (approx.)	Effect	Certainty	Reasons
Wellbeing (positive construct)	4	3,200	$g = 0.10$ (0.01–0.18)	Low	RoB; indirectness; imprecision
Stress or distress	6–8	4,000	$g = 0.26$ (0.10–0.43)	Low	RoB; $I^2 \sim 65\%$; mixed samples
Retention, one-way scheduled	6	3,000+	Median 85–93%	Moderate	Consistent across countries
Retention, two-way bots	2–3	1,700+	42–66%	Low	Few trials; mixed objects
One-way vs two-way outcomes	10	5,225	No clean winner	Very low	Underpowered split
LMIC vs HIC outcomes	2 LMIC	~1,100	Similar averages	Very low	Two trials

Overall: low certainty that messaging programmes produce a small short-term benefit on stress or distress; low certainty that they move wellbeing as a positive construct; moderate certainty that one-way scheduled messages hold about eight in ten starters through a short course.

07

Discussion

PULL QUOTE

We found no completed adult Indian randomised trial that delivers a wellbeing or stress programme primarily by SMS or WhatsApp.

What the number means

A Hedges g of 0.18 is a small effect. In the language of a workplace programme it is not nothing and it is not a transformation. If you line people up by wellbeing and shift everyone by two-tenths of a standard deviation, some will notice a slightly easier Tuesday. Most will not write home.

Compare it with the neighbours. Head's 2013 health-behaviour pool was $d = 0.33$. Senanayake's 2019 depression pool was 0.23 and the interval included zero. Smartphone-app meta-analyses for stress often land near $g = 0.27$ before

trim-and-fill, then shrink. Messaging for wellbeing, on the present evidence, sits at the low end of that family. That is the honest placement.

The prediction interval is the sentence that should hang on the product wall. Minus 0.10 to plus 0.46. A new Indian workplace programme could print a null. It could also print a small, useful lift. The data do not license a forecast of $g = 0.40$. Anyone who writes that forecast is selling, not reading.

Machine-learning people have a name for this. Overfitting. Two small clinical trials with twice-daily SMS and a change-score d of 0.67 to 0.85 will seduce a model. The 1,021-person workplace null will not. Regularise. Keep the large nulls in the pool. That is how you stop a story from becoming a pitch.

Check the belief "messages fix wellbeing" against the data, the way you would check a worried thought against what actually happened last week. The belief is too big. The data support a smaller one: "counted messages sometimes reduce distress a little, especially when the sample is already distressed, and they often fail to lift wellbeing as a positive construct."

Two things are true at once. People stay. Effects are small. Hold both.

So stop waiting for a large g before you ship a humane cadence. Ship the cadence. Measure. Update.

Go back to the three phones. On one the message is swiped away. On one it sinks under the group chats. On one it becomes a morning rep. That is not a reason to stop sending. It is a reason to know who is still reading. One-way scheduled SMS in Brazil held 87%. A WhatsApp conversational agent in the United States held 42%. Know the thread before you write into it.

Limitations

This review was conducted as a structured multi-source search by a small team on a single calendar day, 22 September 2026. We did not export every database hit into a shared reference manager with dual independent screening of 3,860 records. PRISMA flow counts are therefore reconstructed from the search we actually ran, not from a librarian audit trail. That lowers confidence in the denominator. It does not invent the ten trials.

Effect-size conversions are imperfect. An adjusted odds ratio of 1.57 is not the same object as a change-score Cohen's d of 0.67. We converted, we flagged, we ran leave-one-out. Residual error remains.

Several wellbeing instruments are trademarked in some jurisdictions. We described constructs and did not lean on brand names in the public-facing boxes. The underlying papers used the instruments they used. Readers who need the exact scale should open the primary source.

We excluded pre–post population programmes. That decision is correct for pooling. It throws away reach data. Agyapong's provincial programmes enrolled tens of thousands. They cannot be averaged with a 54-person pilot. They can be cited as evidence that people will subscribe.

Clinical and universal samples share a pool. That inflates heterogeneity. We report the split. We do not pretend the split is a new trial.

No Indian adult trial met eligibility. Absence of evidence is not evidence of absence of effect in India. It is evidence of an evidence gap.

Follow-up is short. Scazufca's contrast was gone by five months. Maciejewski's stress lift was gone at one month. Gold's affect lift faded. A 12-week workplace programme that prints $g = 0.18$ at week eight and $g = 0.00$ at month six is a different product from a durable practice. We cannot grade durability. The data are not there.

WHAT THIS MEANS

What this means for an Indian workplace programme

A supervisor in Nagpur clears the family group and the team group before her first call. The wellness app sits unopened. Do not make her download another app to feel slightly less tired. Put the rep on the thread she already opens.

Count the reps. A programme without a dose is a pep talk. Forty-eight messages in six weeks, as in the Brazilian trial, is a dose. Three gratitude texts a week for three weeks, as in the American healthcare trial, is a dose. "We will send supportive content" is not.

One-way is not second-class. Two-way conversation did not outperform push in this pool. One-way held more people. In a country of 23 languages and uneven bandwidth, a short scheduled message in the language of the home is a more honest first product than a chatbot that wants a paragraph back.

Do not promise flourishing. Promise a small, measured shift in a hard week, and a channel that is still there on Friday. The Japanese workplace trial of 1,021 people is your cautionary tale. Acceptable. Appropriate. Null on distress. Build for that possibility. Instrument the programme so a null is visible, not hidden.

Keep the data in India. Hand off to a human when the message is not enough. That hand-off is coaching, not a diagnosis. The eight-second down-shift, the thank-you sent rather than felt, the three good things typed into a thread — these are practices. They sit with other small drills: writing a thought down to look at it, doing the opposite of what a bad mood asks. They do not belong on a prescription pad.

The India gap is the operational point. We found no completed adult Indian randomised trial that delivers a wellbeing or stress programme primarily by SMS or WhatsApp. Until that trial exists, an HR head in Pune who switches on a WhatsApp programme is running a study, whether she calls it one or not. Treat it as one.

Research gaps, naming the India gap explicitly

The India gap is not a rhetorical flourish. CTRI records that look close fail on channel or on outcome. A Kerala programme used telephone calls, not messaging. A Punjab WhatsApp trial targeted hypertension visit attendance, not wellbeing. Community programmes in Gujarat and Andhra Pradesh used digital tools but were not messaging-primary. A small mixed-methods adaptation of a World Health Organization stress programme for community health workers used an app, with WhatsApp as technical support. Qualitative work from urban India finds older adults willing to receive self-help on WhatsApp. Willingness is not an effect size.

What India still lacks, as of 22 September 2026:

- An adult randomised trial of a WhatsApp- or SMS-delivered wellbeing programme in working-age professionals, with a positive wellbeing index as a primary outcome.
- A pre-specified contrast of one-way daily reps versus two-way coaching-lite, in at least two Indian languages.
- Retention reported as survey completion and as message-open or reply rate, not only as "we sent 48 texts."
- Follow-up at six months, not only at the last scheduled ping.
- A workplace sample that is not a university and not a clinic — the people EmotionApp actually serves.

Until those five exist, the pooled g of 0.18 is a prior, not a promise.

Other gaps are global. Durability is unknown. Mechanisms are untested. We do not know whether a gratitude prompt, a planning cue, a behavioural-activation audio or a supportive sentence is the active ingredient. Bendtsen said so about his own programme. We still cannot say. Frequency as a moderator remains a Head (2013) finding for behaviour, not a finding for wellbeing. Language effects are almost untested outside English, Portuguese and Japanese.

08

FAQ

Q1 Does a daily WhatsApp message improve wellbeing?

A little, on average, and not reliably. Across ten trials the pooled effect was $g = 0.18$. When we kept only wellbeing indices the figure fell to $g = 0.10$. Plan for a small lift or a null. Do not plan for a transformation.

Q2 Will people keep reading the messages?

About eight in ten finish a short one-way course. One-way scheduled SMS and WhatsApp broadcasts held 85–93% in the better-reported trials. A two-way conversational agent held 42%. Retention is the strong number in this paper.

Q3 Is a chatbot better than a simple daily text?

Not on present data. Two-way formats pooled at $g = 0.15$; one-way at $g = 0.22$. Conversation did not win. It did lose more people. Start with a counted one-way cadence unless you have a reason to chat.

Q4 Has this been tested with Indian professionals?

No. We found zero completed adult Indian randomised trials of SMS- or WhatsApp-delivered wellbeing or stress programmes as the primary channel. That is the gap this paper exists to name.

Q5 Is this a substitute for seeing someone?

No. It is a practice delivered as a rep. When the message is not enough, hand off to a human coach. Messaging keeps people in the room. It does not replace the room.

09

References

- Agyapong, V. I. O., Ahern, S., McLoughlin, D. M., & Farren, C. K. (2012). Supportive text messaging for depression and comorbid alcohol use disorder: Single-blind randomised trial. *Journal of Affective Disorders*, 141(2–3), 168–176. <https://doi.org/10.1016/j.jad.2012.02.040>
- Agyapong, V. I. O., Juhás, M., Ohinmaa, A., Omeje, J., Mrklas, K., Suen, V. Y. M., Dursun, S. M., & Greenshaw, A. J. (2017). Randomized controlled pilot trial of supportive text messages for patients with depression. *BMC Psychiatry*, 17, 286. <https://doi.org/10.1186/s12888-017-1448-2>
- Agyapong, V. I. O., Hrabok, M., Vuong, W., Shalaby, R., Noble, J. M., Gusnowski, A., Mrklas, K. J., Li, D., Urchuk, L., Snaterse, M., Surood, S., Cao, B., Li, X.-M., Greiner, R., & Greenshaw, A. J. (2021). Text4Hope: Receiving daily supportive text messages for 3 months during the COVID-19 pandemic reduces stress, anxiety, and depression. *Disaster Medicine and Public Health Preparedness*, 16(4), 1326–1330. <https://doi.org/10.1017/dmp.2021.27>
- Agyapong, B., Shalaby, R., Vuong, W., Gusnowski, A., Surood, S., Greenshaw, A. J., Wei, Y., & Agyapong, V. I. O. (2023). Text4Hope effectiveness in reducing psychological symptoms among young adults in Canada: Longitudinal and naturalistic controlled program evaluation. *Journal of Clinical Medicine*, 12(5), 1942. <https://doi.org/10.3390/jcm12051942>
- Bendtsen, M., Müssener, U., Linderroth, C., & Thomas, K. (2020). A mobile health intervention for mental health promotion among university students: Randomized controlled trial. *JMIR mHealth and uHealth*, 8(3), e17208. <https://doi.org/10.2196/17208>
- Foran, H. M., et al. (2024). An automated conversational agent self-help program: Randomized controlled trial. *Journal of Medical Internet Research*, 26, e53829. <https://doi.org/10.2196/53829>
- Gold, K. J., Dobson, M. L., & Sen, A. (2023). "Three Good Things" digital intervention among health care workers: A randomized controlled trial. *Annals of Family Medicine*, 21(3), 220–226. <https://doi.org/10.1370/afm.2963>
- Goldberg, S. B., Lam, S. U., Simonsson, O., Torous, J., & Sun, S. (2022). Mobile phone-based interventions for mental health: A systematic meta-review of 14 meta-analyses of randomized controlled trials. *PLOS Digital Health*, 1(1), e0000002. <https://doi.org/10.1371/journal.pdig.0000002>
- Head, K. J., Noar, S. M., Iannarino, N. T., & Harrington, N. G. (2013). Efficacy of text messaging-based interventions for health promotion: A meta-analysis. *Social Science & Medicine*, 97, 41–48. <https://doi.org/10.1016/j.socscimed.2013.08.003>
- Li, A. C. M., Wong, K. K. L., Chio, F. H. N., Mak, W. W. S., & Poon, L. W. H. (2022). Delivering mindfulness-based interventions for insomnia, pain, and dysfunctional eating through a text messaging app: Three randomized controlled trials. *Journal of Medical Internet Research*, 24(5), e30073. <https://doi.org/10.2196/30073>
- Linardon, J., Torous, J., Firth, J., Cuijpers, P., Messer, M., & Fuller-Tyszkiewicz, M. (2024). Current evidence on the efficacy of mental health smartphone apps for symptoms of depression and anxiety: A meta-analysis of 176 randomized controlled trials. *World Psychiatry*, 23(1), 146–158. <https://doi.org/10.1002/wps.21183>
- Maciejewski, J., & Smoktunowicz, E. (2023). Low-effort internet intervention to reduce students' stress delivered with Meta's Messenger chatbot (Stressbot): A randomized controlled trial. *Internet Interventions*, 33, 100653. <https://doi.org/10.1016/j.invent.2023.100653>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>

- Sasaki, N., Ogawa, S., Sawada, U., Shimazu, T., Powell, B. J., Takeno, H., Tsutsumi, A., & Imamura, K. (2024). Effectiveness of an online text-based stress management program for employees who work in micro- and small-sized enterprises: A randomized controlled trial. *Internet Interventions*, 37, 100754. <https://doi.org/10.1016/j.invent.2024.100754>
- Sczufca, M., Nakamura, C. A., Seward, N., Didone, T. V. N., Moretti, F. A., da Costa, M. O., et al. (2024). Self-help mobile messaging intervention for depression among older adults in resource-limited settings: A randomized controlled trial. *Nature Medicine*, 30, 1880–1888. <https://doi.org/10.1038/s41591-024-02864-4>
- Schenkel, K., Haug, S., Paz Castro, R., Lüscher, J., Scholz, U., Schaub, M. P., & Radtke, T. (2022). One SMS a day keeps the stress away? A just-in-time planning intervention to reduce occupational stress among apprentices. *Applied Psychology: Health and Well-Being*, 14(4), 1389–1407. <https://doi.org/10.1111/aphw.12340>
- Senanayake, B., Wickramasinghe, S. I., Chatfield, M. D., Hansen, J., Edirippulige, S., & Smith, A. C. (2019). Effectiveness of text messaging interventions for the management of depression: A systematic review and meta-analysis. *Journal of Telemedicine and Telecare*, 25(9), 513–523. <https://doi.org/10.1177/1357633X19875852>
- Skeen, S., Marlow, M., du Toit, S., Melendez-Torres, G. J., Mudekanye, L., Mapalala, E., et al. (2023). Using WhatsApp support groups to promote responsive caregiving, caregiver mental health and child development in the COVID-19 era: A randomised controlled trial of a fully digital parenting intervention. *Digital Health*, 9. <https://doi.org/10.1177/20552076231203893>

About the author

ABOUT THE AUTHOR

Dr Atul Aswani; Mumbai Psychiatrist and trained as a Results Coach

Atul Aswani is a Mumbai Psychiatrist and trained as a Results Coach. He builds EmotionApp, a non-clinical emotional-fitness platform for Indian professionals: counted daily reps, an AI coach with human hand-off, WhatsApp-first delivery, 23 Indian languages, India-resident data, DPDP-compliant, designed under ISO 9001 and ISO 13485. He lives and works in Mumbai.

HOW TO CITE

Aswani A. Messaging as the Medium: A Meta-Analysis of SMS- and WhatsApp-Delivered Wellbeing Interventions. EmotionApp Research Paper 42. Mumbai: Tranquilo Meditek Sytems Private Limited; 2026. <https://emotionapp.in/research/sms-whatsapp-wellbeing-meta-analysis>

LAST UPDATED

Last updated: 22 September 2026

NOTE

EmotionApp is a non-clinical emotional-fitness coaching platform. This paper describes practices, techniques and coaching. It is not therapy, treatment, diagnosis or medical advice.

© 2026 Tranquilo Meditek Sytems Private Limited, Mumbai
· emotionapp.in

SAFETY NOTE

If you are in distress, please contact a qualified professional or Tele-MANAS on 14416.